WHAT IS CLAIMED IS:

1. A computer readable medium including instructions readable by a computer which, when implemented, cause the computer to resolve an overlapping ambiguity string in an input sentence of an unsegmented language by performing steps comprising:

segmenting the sentence into two possible segmentations;

recognizing the overlapping ambiguity string in the input sentence as a function of the two segmentations; and

selecting one of the two segmentations as a function of probability information for the two segmentations.

2. The computer readable medium of claim 1 and further comprising obtaining the probability information from a lexical knowledge base.

3. The computer readable medium of claim 2 wherein the lexical knowledge base comprises a trigram model.

4. The computer readable medium of claim 2 wherein selecting one of the two segmentations comprises classifying the probability information.

5. The computer readable medium of claim 4 wherein classifying comprises Naïve Bayesian Classification.

6. The computer readable medium of claim 1 wherein segmenting the sentence comprises performing a Forward Maximum Matching (FMM) segmentation of the input sentence and a Backward Maximum Matching (BMM) segmentation of the input sentence.

7. The computer readable medium of claim 6 wherein recognizing the overlapping ambiguity string comprises recognizing a segmentation O_f of the overlapping ambiguity string from the FMM segmentation and a segmentation O_b of the overlapping ambiguity string from the BMM segmentation.

8. The computer readable medium of claim 7 wherein selecting one of the two segmentations is a function of a set of context features associated with the overlapping ambiguity string.

9. The computer readable medium of claim 8 wherein the set of context features comprises words around the overlapping ambiguity string.

10. The computer readable medium of claim 8 wherein selecting one of the two segmentations comprises classifying the probability information of the set of context features and O_f .

11. The computer readable medium of claim 10 wherein selecting one of the two segmentations comprises classifying the probability information of the set of context features and O_b.

12. The computer readable medium of claim 8 wherein selecting comprising determining which of O_f or O_b has a higher probability as a function of the set of context features.

13. The computer readable medium of claim 1 wherein the unsegmented language is Chinese.

14. A method of segmentation of a sentence of an unsegmented language, the sentence having an overlapping ambiguity string (OAS), the method comprising the steps of:

generating a Forward Maximum Matching (FMM) segmentation of the sentence;

generating a Backward Maximum Matching (BMM) segmentation of the sentence;

recognizing an OAS as a function of the FMM and the BMM segmentations; and

selecting one of the FMM segmentation and the BMM segmentation as a function of probability information.

15. The method of claim 14 wherein the step of selecting includes determining a probability

associated with each of the FMM segmentation of the overlapping ambiguity string and the BMM segmentation of the overlapping ambiguity string, the G scores comprising probability information.

16. The method of claim 15 wherein determining the probability information comprises using an N-gram model.

17. The method of claim 16 wherein determining the probability comprises using probability information about a first word of the overlapping ambiguity string.

18. The method of claim 17 wherein determining the probability comprises using probability information about a last word of the overlapping ambiguity string.

19. The method of claim 16 wherein using the N-gram model comprises using information about context words around the overlapping ambiguity string.

20. The method of claim 16 wherein the N-gram model comprises using information about a string of words comprising a first word of the overlapping ambiguity string and two context words to the left of the first word.

21. The method of claim 20 wherein the N-gram model comprises using information about a string of words comprising a last word of the overlapping ambiguity string and two context words to the right of the last word.

22. The method of claim 15 wherein selecting includes using Naïve Bayesian Classifiers.

23. The method of claim 14 and further comprising receiving information from a lexical knowledge base comprising a trigram model.

24. The method of claim 23 and further comprising receiving an ensemble of Naïve Bayesian Classifiers.

25. A method of constructing information to resolve overlapping ambiguity strings in an unsegmented language comprising the steps of:

recognizing overlapping ambiguity strings in a training data;

replacing the overlapping ambiguity strings with tokens;

generating an N-gram language model comprising information on constituent words of the overlapping ambiguity strings.

26. The method of claim 25 wherein generating the N-gram language model comprises generating a trigram model.

27. The method of claim 25 and further comprising generating an ensemble of classifiers as a function of the N-gram model.

28. The method of claim 25 wherein recognizing the overlapping ambiguity strings comprises:

generating a Forward Maximum Matching (FMM) segmentation of each sentence in the training data;

generating a Backward Maximum Matching (BMM) segmentation of each sentence in the training data;

recognizing an OAS as a function of the FMM and the BMM segmentations of each sentence in the training data.

29. The method of claim 28 and further comprising generating an ensemble of classifiers as a function of the N-gram model.

30. The method of claim 29 wherein generating the ensemble of classifiers includes approximating an a probability of a segmentation of each overlapping ambiguity string as being equal to the product of individual unigram probabilities of individual words in the segmentation of the overlapping ambiguity string.

31. The method of claim 30 wherein generating the ensemble of classifiers includes approximating a joint probability of a set of context features conditioned on an existence of the segmentation of each overlapping ambiguity string as a function of a corresponding probability of a leftmost and a rightmost word of the corresponding overlapping ambiguity string.